Applications of Stationary Anonymous Sequential Games to Multiple Access Control in Wireless Communications

Eitan Altman
INRIA, 2004 Route des Lucioles, P.B. 93,
06902 Sophia Antipolis Cedex, France
email: Eitan.Altman@inria.fr
URL: http://www-sop.inria.fr/members/Eitan.Altman/

Piotr Więcek
Institute of Mathematics and Computer Science,
Wrocław University of Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland,
Tel.: +48-071-320-31-60 Fax: +48-071-328-07-51
email: Piotr.Wiecek@pwr.wroc.pl

***Abstract*—We consider in this paper dynamic Multiple Access (MAC) games between a random number of players competing over collision channels. Each of several mobiles involved in an interaction determines whether to transmit at a high or at a low power. High power decreases the lifetime of the battery but results in smaller collision probability. We formulate this game as an anonymous sequential game with undiscounted reward which we recently introduced and which combines features from both population games (infinitely many players) and stochastic games. We briefly present this class of games and basic equilibrium existence results for the total expected reward as well as for the expected average reward. We then apply the theory in the MAC game.**

Keywords: Stochastic game, Population game, Anonymous sequential game, Average reward, Total reward, Stationary policy

I. Introduction

A well known class of games that involves a continuum of atomless players are evolutionary games, in which pairs of players that play a matrix game are selected at random, see [1]. This game allows to predict the fraction of the population (or of populations in the case of several classes) that play each possible action at equilibrium. This modelling paradigm along with the solution concept called Evolutionary Stable Strategy has had quite a success in biology (we refer the reader to [2], [3]). The player's type in these games is fixed, and the actions of the players determine directly their utilities.

An extension of this model is needed to model the possibility that the player's class may change randomly in time, and to allow the utility of a player to depend not only on the current actions of players but also on future interactions. The class of the player is called its individual or private state. The choice of an action by a player should then take into account not only the game played at the present individual state but the future state evolution. We are interested in particular in the case where the action of a player not only impacts the current utility but also the transition probabilities to the next state.

We use the framework of anonymous sequential games, introduced by B. Jovanovic and R.W. Rosenthal in 1988 in [4]. In that work, each player's utility is given as the expected discounted utility over an infinite horizon. The theory of anonymous sequential games with discounted utilities was further developped in [5], [6], [7], [8], [9]. We build in this paper on a recent theoretical work [13], [15] that establishes the theory of anonymous sequential games for both the cases of total expected cost criterion as well as the time average expected cost. Other applications to power control, to road traffic and to maintenance have been reported in [14] and in [13], [15]. Similar extensions have been proposed and studied for the framework of evolutionary games in [10], [11]. The analysis there turns out to be simpler since the utility in each encounter between two players turns out to be bilinear there.

The structure of the paper is as follows. We begin with a section with an overview of the model. The following section presents the theory for the average and the total reward. We then apply this framework to the dynamic MAC game and derive the equilibrium policy.

II. Anonymous sequential games

The presentation below is based on [13]. The anonymous sequential game is described by the following objects:

- We assume that the game is played in discrete time, that is $t \in \{1, 2, \ldots\}$.
- The game is played by an infinite number (continuum) of players. Each player has his own private state $s \in S$, changing over time. We assume that S is a finite set.
- The global state, μ^t, of the system at time t, is a probability distribution over S. It describes the proportion of the population, which is at time t in each of the individual states. We assume that each player has an ability to observe the global state of the game, so from his point of view the state of the game at time t is[1] $(s_t, \mu^t) \in S \times \Delta(S)$.
- The set of actions available to a player in state (s, μ) is a nonempty set $\mathsf{A}(s, \mu)$, with $A := \bigcup_{(s,\mu)\in S\times\Delta(S)} \mathsf{A}(s, \mu)$

[1]Here and in the sequel for any set B, $\Delta(B)$ denotes the set of all the finite-support probability measures on B. In particular, if B is a finite set, it denotes the set of all the probability measures over B. In such a case we always assume that $\Delta(B)$ is endowed with Euclidean topology.

– a finite set. We assume that the mapping A is an upper semicontinuous function.

- Global distribution of the state-action pairs at any time t is given by the measure $\tau^t \in \Delta(S \times A)$. The global state of the system μ^t is the marginal of τ^t on S.
- An individual's immediate reward at any stage t, when his private state is s_t, he plays action a_t and the global state-action measure is τ^t is $u(s_t, a_t, \tau^t)$. It is a (jointly) continuous function.
- The transitions are defined for each individual separately with the transition function $Q : S \times A \times \Delta(S \times A) \to \Delta(S)$ which is also a (jointly) continuous function. We will write $Q(\cdot|s_t, a_t, \tau^t)$ for the distribution of the individual state at time $t+1$, given his state at time t, s_t, his action a_t and the state-action distribution of all the players.
- The global state at time $t+1$ will be given by[2] $\Phi(\cdot|\tau^t) = \sum_{s\in S}\sum_{a\in A} Q(\cdot|s, a, \tau^t)\tau^t_{sa}$.

Any function $f : S \times \Delta(S) \to \Delta(A)$ satisfying $\text{supp} f(s,\mu) \subset \mathsf{A}(s,\mu)$ for every $s \in S$ and $\mu \in \Delta(S)$ is called a *stationary policy*. We denote the set of stationary policies in our game by $\mathcal{U}$.

A. Average reward

We define the *long-time average reward* of a player using stationary policy f when all the other players use policy g and the initial state distribution (both of the player and his opponents) is μ^1, to be

$$J(\mu^1, f, g) = \limsup_{T\to\infty} \frac{1}{T} E^{\mu^1, Q, f, g} \sum_{t=1}^{T} u(s_t, a_t, \tau^t).$$

Further, we define a stationary strategy f and a measure $\mu \in \Delta(S)$ to be an equilibrium in the long-time average reward game if for every other stationary strategy $g \in \mathcal{U}$,

$$J(\mu, f, f) \geq J(\mu, g, f)$$

and, if $\mu^1 = \mu$ and all the players use policy f then $\mu^t = \mu$ for every $t \geq 1$.

B. Total reward

To define the total reward in our game let us distinguish one state in S, say s_0 and assume that $\mathsf{A}(\mathsf{s}_0, \mu) = \{a_0\}$ independently of μ for some fixed a_0. Then the *total reward* of a player using stationary policy f when all the other players apply policy g and the initial distribution of the states of his opponents is μ^1, while his own is ρ^1, is defined in the following way:

$$\overline{J}(\rho^1, \mu^1, f, g) = E^{\rho^1, \mu^1, Q, f, g} \sum_{t=1}^{\mathcal{T}-1} u(s_t, a_t, \tau^t),$$

where $\mathcal{T}$ is the moment of the first arrival of the process s_t to s_0. We interpret it as the reward accumulated by the player over whole of his lifetime. State s_0 is an artificial state (so is action a_0) denoting that a player is dead. μ^1 is the distribution of the states across the population when he is born, while ρ^1 is the distribution of initial states of new-born players. The fact that after some time the state of a player can become again different from s_0 should be interpreted as that after some time the player is replaced by some new-born one.

The notion of equilibrium for the total reward case will be slightly different from that for the average reward. We define a stationary strategy f and a measure $\mu \in \Delta(S)$ to be in equilibrium in the total reward game if for every other stationary strategy $g \in \mathcal{U}$,

$$\overline{J}(\rho, \mu, f, f) \geq \overline{J}(\rho, \mu, g, f),$$

where $\rho = Q(\cdot|s_0, a_0, \tau(f, \mu))$ and $(\tau(f,\mu))_{sa} = \mu_s(f(s))_a$ for all $s \in S$, $a \in A$, and, if $\mu^1 = \mu$ and all the players use policy f then $\mu^t = \mu$ for every $t \geq 1$.

III. Existence of the Stationary Equilibrium

Average-reward Case

We next introduce an important assumption on the individual state process.

(A1) The set of individual states of any player S can be partitioned into two sets S_0 and S_1 such that for every state-action distribution of all the other players $\tau \in \Delta(S \times A)$:

(a) All the states from S_0 are transient in the Markov chain of individual states of a player using any $f \in \mathcal{U}$.

(b) The set S_1 is strongly communicating.

In [13], [15], a couple of equivalent definitions of "strongly communicating" are cited. It is shown in [15] through an example that without assumption (A1) the average-reward anonymous sequential game may have no stationary equilibria at all.

Theorem 1: Every anonymous sequential game with long-time average payoff satisfying (A1) has a stationary equilibrium.

The Total-reward Case

We will assume the following:

(T1) There exists a $p_0 > 0$ such that for any fixed state-action measure τ and under any stationary policy f the probability of getting from any state $s \in S \setminus \{s_0\}$ to s_0 in $|S| - 1$ steps is not smaller than p_0.

Remark 1: The total reward model, specifically when (T1) is assumed, bears a lot of resemblance to an exponentially discounted model where the discount factor is allowed to fluctuate over time, which suggests that the results in the two models should not differ much. Note however that there is one essential difference between these two models. The 'discount factor' in the total reward model (which is the ratio of those who stay alive after a given period to those who were alive at its beginning) appears not only in the cumulative reward of the players but also in the stationary state of the game, and thus also in the per-period rewards of the players. Thus this is an essentially different (and slightly more complex) problem. On the other hand, the fact that each of the players lives for

[2] Note that its transition is deterministic.

a finite period and then is replaced by another player, with a fixed fraction of players dead and fixed fractions of players in each of the states when the game is in a stationary state, makes this model similiar to the average reward one. In fact, using the renewal theorem, we can relate the rewards of the players in the total reward model with those in the respective average reward model. This relation in used a couple of times in our proofs.

Theorem 2: Every anonymous sequential game with total reward satisfying (T1) has a stationary equilibrium.

IV. Application: Medium Access Game

A. The Model

Consider the following MAC (Medium ACcess) game between mobile phones. Time is slotted. At any given time t, a mobile finds itself competing with N_t other mobiles for the access to a channel. N_t is assumed to have Poisson distribution with parameter λ. We shall formulate this as a sequential anonymous game as follows.

- **Individual state** A mobile has three possible states: F (full) AE (Almost Empty) and E (Empty).
- **Actions** There are two actions: transmit at high power H or low power L. At state AE a mobile cannot transmit at high power, while at E it cannot transmit at all.
- **Transition probabilities** From state AE the mobile moves to state E with probability p_E and otherwise remains in AE. At state E the mobile has to recharge. It moves to state F after one time unit. A mobile in state F transmitting with power r moves to state AE with probability proportional to r and given by αr for some constant $\alpha > 0$.
- **Payoff** Consider a given cellular phone that transmits a packet. Assume that x other packets are transmitted with high power and y with low power to the same base station. A packet transmitted with low power is received successfully with some probability q if it is the only packet transmitted, i.e. $y = 0$, $x = 0$. Otherwise it is lost. A packet transmitted with high power is received successfully with some probability $Q > q$ if it is the only packet transmitted at high power, i.e. $x = 0$. The immediate payoff is 1 if the packet is successfully transmitted. It is otherwise zero. In addition there is a constant cost $c > 0$ for recharging the battery. Aggregate utility for a player is then computed as long-time average of the per-period payoffs.

Suppose p is the fraction of population that transmits at high power in state F, and that μ_F, μ_{AE} and μ_E are fractions of players in respective states. Then probability of success for a player transmitting at high power is

$$QP(x=0) = Q(e^{-\lambda} + \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}(1-p\mu_F)^k)$$
$$= Qe^{-\lambda} e^{\lambda(1-p\mu_F)} = Qe^{-\lambda p \mu_F}, \qquad (1)$$

while the probability of success when a player transmits at low power is

$$qP(x+y=0) = q(e^{-\lambda} + \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \mu_E^k)$$
$$= qe^{-\lambda} e^{\lambda \mu_E} = qe^{\lambda(\mu_E - 1)}. \qquad (2)$$

These values do not depend on actual numbers of players applying respective strategies – only on fractions of players in each of the states using different actions. Thus instead of considering an n-player game for any fixed n it is reasonable to apply the anonymous game formulation with $\tau = [\tau_{F,H}, \tau_{F,L}, \tau_{AE,L}, \tau_E]$ denoting the vector of fractions of players in respective states and using respective actions, with immediate rewards

$$u(s,a,\tau) = \begin{cases} Qe^{-\lambda \tau_{F,H}}, & \text{when } a = H \\ qe^{\lambda(\tau_E - 1)}, & \text{when } a = L \\ -c, & \text{when } s = E \end{cases}$$

and transition probabilities defined by matrix

$$\mathbb{Q}(a,\tau) = \begin{bmatrix} 1-\alpha a & \alpha a & 0 \\ 0 & 1-p_E & p_E \\ 1 & 0 & 0 \end{bmatrix}.$$

B. The Solution

The stationary state of the chain of the private states of a player using policy f prescribing him to use high power with probability p when in state F is

$$\frac{1}{\alpha(pH + (1-p)L)(p_E+1) + p_E}$$
$$\times [p_E, \alpha(pH+(1-p)L), p_E\alpha(pH+(1-p)L)].$$

Thus computations yield that his respected long-run average reward is of the form

$$\frac{Ap+B}{Cp+D} \quad \text{with}$$
$$A = p_E Q e^{-\lambda \tau_{F,H}} + ((H-L)\alpha - p_E) q e^{\lambda(\tau_E - 1)}$$
$$-c\alpha p_E (H-L),$$
$$B = (L\alpha + p_E) q e^{\lambda(\tau_E-1)} - c\alpha p_E L,$$
$$C = \alpha(H-L)(p_E+1),$$
$$D = \alpha L(p_e + 1) + p_E.$$

It can be either a strictly increasing, a constant or a strictly decreasing function of p, depending on whether $AD > BC$, $AD = BC$ or $AD < BC$, and thus the best response of a player against the aggregated state-action vector τ is $p = 1$ when $AD > BC$, any $p \in [0,1]$ when $AD = BC$ or $p = 0$ when $AD < BC$. This leads to the following conclusion: since by Theorem 1 this anonymous game has an equilibrium, one of the three following cases holds:

(a) If

$$\Big[p_E Q e^{-\frac{\lambda p_E}{\alpha H(p_E+1)+p_E}}$$
$$+((H-L)\alpha - p_E) q e^{-\frac{\lambda(\alpha H + p_E)}{\alpha H(p_E+1)+p_E}} - a\alpha p_E(H-L)\Big]$$

$$\times\Big[\alpha L(p_E+1)+p_E\Big]$$

$$> \Big[(L\alpha{+}p_E)qe^{-\frac{\lambda(\alpha H+p_E)}{\alpha H(p_E+1)+p_E}}-c\alpha Lp_E\Big]\Big[\alpha(H{-}L)(p_E{+}1)\Big]$$

then all the players use high power in state F at equilibrium.

(b) If

$$\Big[p_EQe^{-\frac{\lambda p_E}{\alpha L(p_E+1)+p_E}}$$

$$+((H-L)\alpha-p_E)qe^{-\frac{\lambda(\alpha L+p_E)}{\alpha L(p_E+1)+p_E}}-a\alpha p_E(H-L)\Big]$$

$$\times\Big[\alpha L(p_E+1)+p_E\Big]$$

$$< \Big[(L\alpha{+}p_E)qe^{-\frac{\lambda(\alpha L+p_E)}{\alpha L(p_E+1)+p_E}}-c\alpha Lp_E\Big]\Big[\alpha(H{-}L)(p_E{+}1)\Big]$$

then all the players use low power in state F at equilibrium.

(c) If none of the above inequalities holds than we need to find p^* satisfying

$$\Big[p_EQe^{-\lambda\tau_{F,H}}+((H{-}L)\alpha{-}p_E)qe^{\lambda(\tau_E-1)}-c\alpha p_E(H{-}L)\Big]$$

$$\times\Big[\alpha L(p_e+1)+p_E\Big]$$

$$= \Big[(L\alpha+p_E)qe^{\lambda(\tau_E-1)}-c\alpha p_EL\Big]\Big[\alpha(H-L)(p_E+1)\Big]$$

with

$$\tau_{F,H}=\frac{p^*p_E}{\alpha(p^*H+(1-p^*)L)(p_E+1)+p_E} \qquad (3)$$

and

$$\tau_E=\frac{\alpha(p^*H+(1-p^*)L)p_E}{\alpha(p^*H+(1-p^*)L)(p_E+1)+p_E}. \qquad (4)$$

Then all the players use policy prescribing to use high power with probability p^* in state F at equilibrium.

Remark 2: It is worth noting here that some generalizations of the model presented above can be considered. We can assume that there are more energy levels and more powers at which players could transmit in our game (similarly as in [14]). We can also assume that the players do not always transmit, only with some positive probability (then the individual state becomes two-dimensional, consisting of player's energy state and an indicator of whether he has something to transmit or not). Both these generalizations are tractable within our framework, though the computations become more involved.

V. Discussion and Conclusions

The anonymous sequential games that we used here have various common elements with the classical traffic assignment problem [12]. Both problems deal with an infinity of players so as to model large populations. In both frameworks, players can be in different individual states. In the traffic assignment problem, a class can be characterized by a source-destination pair, or by a vehicle type (car, pedestrian or bicycle). In contrast, the class of a player in anonymous sequential games can change in time. Transition probabilities that govern this change may depend not only on the individual's state, but also on the fraction of players that are in each individual state and that use different actions. Furthermore, these transitions are controlled by the player. A strategy of a player of a given class in the traffic assignment problem can be identified as the probability it would choose a given action (path) among those available to its class. The definition of a strategy in our case is similar, except that now the probability for choosing different actions should be specified for each possible individual state. The class of anonymous sequential games is a powerful tool for modeling competition between populations. In this paper we illustrated its usefulness in network engineering. There are however many theoretical open questions that require further reseach on this class of games such as existence of Evolutionary Stable Strategy (which is a standard equilibrium notion in population games) and the definition and the convergence of replicator dynamics in that setting.

VI. Acknowledgements

The work of the first author has been partially supported by the European Commission within the framework of the CONGAS project FP7-ICT-2011-8-317672. The work of the second author was supported by the NCN Grant no DEC-2011/03/B/ST1/00325.

References

[1] Maynard Smith, J.: Game Theory and the Evolution of Fighting. In: Maynard Smith, J. (ed.): On Evolution, pp. 8–28. Edinburgh University Press (1972)

[2] Cressman, R.: Evolutionary Dynamics and Extensive Form Games. MIT Press, Cambridge, MA (2003)

[3] Vincent, T.I., Brown, J.S.: Evolutionary Game Theory, Natural Selection and Darwinian Dynamics. Cambridge University Press, New York (2005)

[4] Jovanovic, B., Rosenthal, R.W.: Anonymous Sequential Games. Journal of Mathematical Economics 17, 77–87 (1988)

[5] Bergin, J., Bernhardt, D.: Anonymous sequential games with aggregate uncertainty. Journal of Mathematical Economics 21, 543–562 (1992)

[6] Bergin, J., Bernhardt, D.: Anonymous Sequential Games: Existence and Characterization of Equilibria. Economic Theory 5(3), 461–89 (1995)

[7] Sleet, C.: Markov perfect equilibria in industries with complementarities. Economic Theory 17(2), 371–397 (2001)

[8] Chakrabarti, S.K.: Pure strategy Markov equilibrium in stochastic games with a continuum of players. J. Math. Econom. 39(7), 693ï£¡-724 (2003)

[9] Adlakha, S., Johari, R.: Mean Field Equilibrium in Dynamic Games with Strategic Complementarities. Operations Research, 61(4), 971–989 (2013)

[10] Altman, E., Hayel, Y.: Stochastic Evolutionary Games. Proceedings of the 13th Symposium on Dynamic Games and Applications, Wroclaw, Poland, 30th June–3rd July, (2008)

[11] Altman, E., Hayel, Y., Tembine, H., El-Azouzi, R.: Markov decision Evolutionay Games with Time Average Expected Fitness Criterion. 3rd International Conference on Performance Evaluation Methodologies and Tools, (Valuetools), Athens, Greece, 21–23 October, (2008)

[12] Wardrop, J.G.: Some theoretical aspects of road traffic research. Proc. Inst. Civ. Eng. 2, 325–378 (1952)

[13] Piotr Więcek and Eitan Altman, Stationary Anonymous Sequential Games with Undiscounted Rewards, International conference on NETwork Games, COntrol and OPtimization (NETGCOOP) 2011.

[14] Więcek, P., Altman, E., Hayel, Y.: Stochastic State Dependent Population Games in Wireless Communication. IEEE Transactions on Automatic Control 56(3), 492–505 (2011)

[15] Piotr Więcek and Eitan Altman, Stationary Anonymous Sequential Games with Undiscounted Rewards, Archived at hal-00947313, http://hal.inria.fr/hal-00947313